

**Request to Archive
With The National Centers for Environmental Information
For National Weather Service (NWS) Upper Air Observations Data (DSI 6212)
Provided by NWS FSO**

2013-12-02

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Jeff Arnfield
NCDC/GCAD/PDB
828-271-4444
Jeff.Arnfield@noaa.gov

2. Name the organization or group responsible for creating the dataset.

NCDC's Global Climate Applications Division (GCAD) produces this dataset from raw inputs received from NWS upper air sites

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

This dataset, previously known as DSI-6213, contains individual upper air flight data from National Weather Service (NWS) sites using Radiosonde Replacement System (RRS) equipment. These roughly 90 sites are currently US-only, spanning about 7 to 72 degrees latitude and -171 to 172 degrees longitude. Flights typically last up to

Data are encoded at each site using the WMO's BUFR file format, then packaged with brief ASCII formatted flight data and transmitted to NCDC. NCDC has software to decode the files, but such decoding is not part of NCDC's operational processing.

Each BUFR file contains seven separate sections, often referred to as datasets: an administrative/metadata section, five data sections with approximately a one-second-resolution, and a final

NC002019 - an administrative section with 96 elements, including flight metadata

NC002020 - Raw Pressure, Temperature, Humidity (PTU) data, with 36 data elements at each level

NC002021 - Raw GPS "unsmoothed radiosonde" data, with 38 data elements at each level

NC002022 - Raw GPS "smoothed wind" data, with 38 data elements at each level

NC002023 - Processed PTU data, with 47 data elements at each level

NC002024 - Processed u & v winds and position data, with 37 data elements at each level

NC002025 - Mandatory, Significant and special PTU and wind levels data, derived from NC002023 and NC002024, with 33 data elements per level

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)

From 2005

Ongoing as continuous updates to the data record

5. Edition or version number(s) of the dataset:

Dataset had not been formally versioned in the past. Arbitrarily version 1.2

6. Describe the level to which the data are processed. For example, are these unprocessed raw observations, derived parameters, quality controlled or inter-calibrated data, etc.?

These raw, undecoded observation undergo minimal validation and QC processing. They are validated for readability of the flight metadata portion of the file and flight dates are compared with those in the file name of data package in which they are received. The file is renamed and packaged for monthly archive. No additional processing or validation is performed.

7. Approximate date when the dataset was or will be released to the public:

2005

8. Who are the expected users of the archived data? How will the archived data be used?

various .EDU, .GOV and private companies. There is a growing interest in high vertical resolution radiosonde data.

9. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?

No

10. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?

Flight data files also include a separate header (H) and thermal level (T) file pair for each flight. These files are at a much lower vertical resolution, and are processed and archived each month as the Upper Air US Rawinsonde dataset, previously known as DSI-6301.

NWS's earlier MicroART upper air equipment generated six-second-resolution data, as well as a separate pair of ASCII files for each flight. The monthly data packages are acquired at NCDC by iRODS and submitted for archive as DSI-6211. Only one such site is currently operational. NCDC separately processes and archives decoded versions of the high high resolution data as DSI-621 and DSI-9948.

11. List the input datasets and ancillary information used to produce the data.

Individual flight data packages are received from individual sites via push to NCDC's FTP, which are acquired and submitted for archive by the iRODS system under DSI-6214. Note that files in these data packages are also the basis for the Upper Air US Rawinsonde dataset, previously known as DSI-6301. Reference and procedural information may be consulted during daily and monthly processing.

12. List web pages and other links that provide information on the data.

While there is NCDC Dataset Documentation, it appears no FGDC-compliant metadata record exists for this dataset. The WMO BUFR format is documented, and includes code values that may appear in the data. The files themselves contain a metadata section. Separate station reference files, with minimal information about each station, are included.

13. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.

1. NCDC dataset documentation TD6213, last revised 2008-08-13, describes the file contents and code values.

14. Indicate the data file format(s).

1. BUFR

15. Are the data files compressed?

No

16. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).

Currently, each month a single tar file is generated for each day of the month, containing all BUFR flight files for flights conducted that day.

Individual BUFR files are named as <wban#>_yyyymmddhh, where <wban#> is the site's WBAN number and yyyymmddhh is the nominal time of the flight.

The archived daily .tar files are named as 6213yyyymmdd.tar, where yyyymmdd is the year/month/day for the BUFR files contained.

17. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?

Sample data packages, as submitted for archive, may be accessed at humboldt:proc_sm/rrs/data/archive.

18. What is the total data volume to be submitted?

Historic Data: all historic data or data submitted as a completed collection.

Total Data Volume: 100GB

Number of Data Files: 2900

Continuous Data: data volume rate for a continuous data production.

Total Data Volume Rate: 1.6GB per Month

Data File Frequency: 31 per Month

Data Production Start: 2005-08

19. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

Occasionally late data are received. While the individual, raw flight data should still reach the archive following the iRODS acquisition process, it would require resubmission to ensure the final dataset is up-to-date.

Separately, the most common use case for the data are to request a specific flight or flight range for one or perhaps several sites. In the future, it may make more sense to archive data by station/month, as we do for the US Rawinsonde set, rather than a day at a time

20. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location: All US states and most territories

System Name: NWS upper air sites

System Owner: NWS

Additional Information: individual sites push data to NCD

21. What are the possible methods for submitting the data to NCEI? Select all that apply.

1. FTP PUSH

22. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

1. Direct download links

23. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

No known constraints apply to the data.

24. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.

We have been generating and archive these NOAA-provided data since 2005. Good stewardship and continued credibility as a data center are most compelling benefits

25. Are the data archived at another facility or are there plans to do so? Please explain.

No

26. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?

We have been acquiring and archiving such data from NWS for many years. This dataset has itself been archived here since 2005.

27. Do you have a data management plan for your data?

No

28. Have funds been allocated to archive the data at NCEI?

No

29. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.

N/A

30. Is there a desired deadline for NCEI to archive and provide access to the data?

Archive by: 2014-03

Accessible by:

31. Add any other pertinent information for this request.

None